Sufficient Estimators [1]

Math Notes | Larry Cui

April 11, 2022

After the expected value and variance of an estimator, we now discuss a third feature: sufficiency. Whether or not an estimator is sufficient refers to the amount of "information" it contains about the unknown parameter θ. We start by defining what an estimator for true θ is sufficient means.

1 Definition and Examples

Definition of Sufficiency

Let $X_1 = k_1, X_2 = k_2, \ldots, X_n = k_n$ be a random sample of size n from $p_X(k;\theta)$, let $\hat{\theta} = h(X_1, X_2, \ldots, X_n)$ be the estimator function for the θ_e, and let $p_{\hat{\theta}}(\theta_e;\theta)$ be the probability from $p_X(k;\theta)$ whatever the domain of $\hat{\theta}(X)$ will take as long as the result value is θ_e.

Let likelihood function, a function about θ, be $L(\theta) = p(k_1)p(k_2)\cdots p(k_n)$.

We say estimator $\hat{\theta}$ is sufficient for θ if,

$$\frac{L(\theta)}{p_{\hat{\theta}}(\theta_e;\theta)} = b(k_1, k_2, \ldots, k_n) \quad \text{or} \quad L(\theta) = p_{\hat{\theta}}(\theta_e;\theta) \cdot b(k_1, k_2, \ldots, k_n)$$

Where $b(k_1, k_2, \ldots, k_n)$ is some constant independent of θ.

The most confusing part of the above definition is the probability $p_{\hat{\theta}}(\theta_e;\theta)$. We use two examples to explain this probability.

Example 1 Suppose a random sample of size n is taken from the Bernoulli pdf,

$$p_X(k;p) = p^k(1-p)^{1-k} \quad , \quad k = 0, 1$$

where p (in definition we use θ to denote parameter, but here we pick p as a convention for Bernoulli distribution) is an unknown parameter for success, i.e., when $k = 1$. We know a maximum likelihood estimator for p is $\hat{p} = \frac{1}{n}\sum_{i=1}^{n} X_i$. To check if this estimator is sufficient, we need the likelihood function and probability of $p_{\hat{\theta}}(\theta_e;\theta)$.

[1] Refer to https://online.stat.psu.edu/stat415/lesson/24

Because every trial is independent, the likelihood function is simply the product of all pdfs:

$$L(\theta) = p^{k_1}(1-p)^{1-k_1}\cdots p^{k_n}(1-p)^{1-k_n} = p^{\sum_{i=1}^n k_i}(1-p)^{n-\sum_{i=1}^n k_i}$$

Let $n\overline{X}$ denote $\sum_{i=1}^n k_i$. We can see that $p_{\hat{\theta}}(\theta_e;\theta)$ in this example is the probability when in a total of n trials there're $n\overline{X}$ successes. And we know that this above sample of size n just gives us one of many situations where the total successful trials equals to $n\overline{X}$:

$$\binom{n}{n\overline{X}} p^{n\overline{X}}(1-p)^{n-n\overline{X}}$$

Following the sufficiency definition, we divide $L(\theta)$ by $p_{\hat{\theta}}(\theta_e;\theta)$, thus in fact we will have a conditional probability of the sample:

$$\frac{L(\theta)}{p_{\hat{\theta}}(\theta_e;\theta)} = \frac{p^{n\overline{X}}(1-p)^{n-n\overline{X}}}{\binom{n}{n\overline{X}}p^{n\overline{X}}(1-p)^{n-n\overline{X}}} = \binom{n}{n\overline{X}}^{-1}$$

The last term $\binom{n}{n\overline{X}}^{-1}$ is not a function of p, so we conclude that estimator $\hat{p}$ is sufficient.

Example 2 Suppose a random sample of size n is taken from the Poisson pdf, $p_X(k) = e^{-\lambda}\lambda^k/k!$. We have an estimator $\hat{\lambda} = \sum_{i=1}^n X_i$.

Likelihood function is,

$$L(\lambda) = \prod_{i=1}^n e^{-\lambda}\lambda^{k_i}/k_i! = e^{-n\lambda}\lambda^{\sum_{i=1}^n k_i}/\prod_{i=1}^n k_i!$$

This example is that during n terms with a $\lambda = n\lambda$, a total of $\sum_{i=1}^n X_i$ cases happen, the probability of this situation is,

$$p_{\hat{\lambda}} = e^{-n\lambda}n\lambda^{\sum_{i=1}^n k_i}/(\sum_{i=1}^n k_i)!$$

If we divide above terms,

$$\frac{L(\lambda)}{p_{\hat{\lambda}}} = \frac{e^{-n\lambda}\lambda^{\sum_{i=1}^n k_i}\cdot(\sum_{i=1}^n k_i)!}{e^{-n\lambda}n\lambda^{\sum_{i=1}^n k_i}\cdot\prod_{i=1}^n k_i!} = \frac{(\sum_{i=1}^n k_i)!}{n^{\sum_{i=1}^n k_i}\prod_{i=1}^n k_i!}$$

This result is a constant, independent of λ, so we conclude the estimator is sufficient.

2 Factorization Criterion/Fisher-Neyman Theorem

From the above two examples, we can see that to find if the estimator is sufficient, the most difficult part is to determine the probability of the estimate. If the estimator function or the pdf itself is complicated, like normal distribution, we may find it increasingly prohibitive to find the probability. Fortunately, we have a factorization theorem to help us.

Fisher-Neyman Theorem Let $X_1 = k_1, X_2 = k_2, \ldots, X_n = k_n$ be a random sample of size n from $p_X(k;\theta)$. The estimator $\hat{\theta} = h(X_1, X_2, \ldots, X_n)$ is sufficient for θ if and only if there are functions $g[h(k_1, k_2, \ldots, k_n);\theta]$ and $b(k_1, k_2, \ldots, k_n)$ such that

$$L(\theta) = g[h(k_1, k_2, \ldots, k_n);\theta] \cdot b(k_1, k_2, \ldots, k_n) \tag{1}$$

where the function $b(k_1, k_2, \ldots, k_n)$ does not involve the parameter θ.

Proof: Let's assume that $L(\theta)$ can be factorized into $g[h(k_1, k_2, \ldots, k_n);\theta]$ and $b(k_1, k_2, \ldots, k_n)$. If we can convert $g[h(k_1, k_2, \ldots, k_n);\theta]$ back to $p_{\hat{\theta}}(\theta_e;\theta) \cdot b'$, where b' is some constant, we prove the sufficiency.

Let c be some end value of estimator function, $c = \theta_e = h(k_1, k_2, \ldots, k_n)$. It's probable that we can have multiple samples of the same size n, each having a different combination of k values, but when put into the function, each having the same result of c. We name the total samples set A,

$$A = \{\{k_1, k_2, \ldots, k_n\}, \{k_1^{'}, k_2^{'}, \ldots, k_n^{'}\}, \{k_1^{''}, k_2^{''}, \ldots, k_n^{''}\}, \ldots\}$$

The total probability of $p_{\hat{\theta}}(c;\theta)$ is simply the sum of each and every sample pdf:

$$\begin{aligned}
p_{\hat{\theta}}(c;\theta) &= \sum_A \prod_{i=1}^{n} p_X(k_i) && \triangleright: \ \textstyle\prod_{i=1}^{n} p_X(k_i) = L(\theta) \\
&= \sum_A g[h(k_1, k_2, \ldots, k_n);\theta] \cdot b(k_1, k_2, \ldots, k_n) && \triangleright: \text{assume equation (1) hold} \\
&= \sum_A g[c;\theta] \cdot b(k_1, k_2, \ldots, k_n) && \triangleright: \text{set A is so constructed that each sample has value } c \\
&= g[c;\theta] \cdot \sum_A b(k_1, k_2, \ldots, k_n)
\end{aligned}$$

How many samples are there in set A is determined by value c and estimator function h, so we are safe to say that $\sum_A b(k_1, k_2, \ldots, k_n)$ is a constant independent of θ. We are done by converting $g[c;\theta]$ to $p_{\hat{\theta}}(c;\theta)$:

$$\begin{aligned}
L(\theta) &= g[h(k_1, k_2, \ldots, k_n);\theta] \cdot b(k_1, k_2, \ldots, k_n) \\
&= p_{\hat{\theta}}(c;\theta) \cdot \frac{b(k_1, k_2, \ldots, k_n)}{\sum_A b(k_1, k_2, \ldots, k_n)}
\end{aligned}$$

Example 3 Let $X_1, X_2, \ldots, X_n$ be a random sample from a normal distribution with mean μ and variance 1. Find a sufficient statistic for the parameter μ.

We can simplify the joint probability by timing pdf of all sample elements:

$$f(x_1, x_2, \ldots, x_n) = \frac{1}{\sqrt{2\pi}} e^{\left[-\frac{1}{2}(x_1-\mu)^2\right]} \cdots \frac{1}{\sqrt{2\pi}} e^{\left[-\frac{1}{2}(x_n-\mu)^2\right]} = \frac{1}{(2\pi)^{n/2}} e^{\left[-\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2\right]} \tag{2}$$

We now factorize the exponential part of equation (2),

$$\begin{aligned}
\exp\left[-\frac{1}{2}\sum_{i=1}^{n}(x_i-\mu)^2\right] &= \exp\left[-\frac{1}{2}\sum_{i=1}^{n}(x_i-\overline{x}+\overline{x}-\mu)^2\right] \\
&= \exp\left[-\frac{1}{2}\sum_{i=1}^{n}[(x_i-\overline{x})^2+2(x_i-\overline{x})(\overline{x}-\mu)+(\overline{x}-\mu)^2]\right] \\
&= \exp\left[-\frac{1}{2}\sum_{i=1}^{n}(x_i-\overline{x})^2-(\overline{x}-\mu)\underbrace{\sum_{i=1}^{n}(x_i-\overline{x})}_{0}-\frac{1}{2}\sum_{i=1}^{n}(\overline{x}-\mu)^2\right] \\
&= \exp\left[-\frac{n}{2}(\overline{x}-\mu)^2\right]\cdot\exp\left[-\frac{1}{2}\sum_{i=1}^{n}(x_i-\overline{x})^2\right]
\end{aligned}$$

On the right-hand side of the equation, we have a function of μ as the first term, and a constant independent of μ as the second term. According to Fisher-Neyman theorem, we conclude the sufficient estimator is $\overline{x}$ or the sum $\sum_{i=1}^{n} X_i$.

Example 4 Let $X_1, X_2, \ldots, X_n$ be a random sample from an exponential distribution with parameter θ, $f(x)=\frac{1}{\theta}e^{-\frac{x}{\theta}}$. Find a sufficient statistic for the parameter θ.

$$\begin{aligned}
L(\theta) &= \frac{1}{\theta}e^{-\frac{x_1}{\theta}}\frac{1}{\theta}e^{-\frac{x_2}{\theta}}\cdots\frac{1}{\theta}e^{-\frac{x_n}{\theta}} \\
&= \frac{1}{\theta^n}\exp\left[-\frac{1}{\theta}\sum_{i=1}^{n}x_i\right]
\end{aligned}$$

This is a function about parameter θ only. But we can still regard it as a factorization of the function and a constant 1. So we conclude it's sufficient if the estimator is the sum of variables.

3 Exponential Form

You may not notice that all examples up to now are coming with a pdf that can be written in exponential form. For such pdf, sample sum is always a sufficient estimator.

Exponential Theorem Let $X_1=k_1, X_2=k_2, \ldots, X_n=k_n$ be a random sample from distribution with a pdf of the exponential form:

$$f(x;\theta)=\exp[K(x)p(\theta)+S(x)+q(\theta)]$$

with

1. $K(x)$ and $S(x)$ being functions only of x
2. $p(\theta)$ and $q(\theta)$ being functions only of the parameter θ
3. The support variable x being free of the parameter θ

The estimator $\sum_{i=1}^{n} K(X_i)$ is sufficient for θ.

Proof We use Fisher theorem to prove it. Starting from joint pdf, we have,

$$\begin{aligned} L(\theta) &= \exp[K(x_1)p(\theta) + S(x_1) + q(\theta)] \cdots \exp[K(x_n)p(\theta) + S(x_n) + q(\theta)] \\ &= \exp\left[p(\theta)\sum_{i=1}^{n} K(X_i) + \sum_{i=1}^{n} S(x_i) + nq(\theta)\right] \\ &= \exp\left[p(\theta)\sum_{i=1}^{n} K(X_i) + nq(\theta)\right] \cdot \exp\left[\sum_{i=1}^{n} S(x_i)\right] \end{aligned}$$

We successfully factorize the likelihood function into two parts, one is a function of θ, the other not. From the above equation, we can see that the $\hat{\theta} = \sum_{i=1}^{n} K(X_i)$ is a sufficient statistic for true θ.

Example 5 Let $X_1, X_2, \ldots, X_n$ be a random sample from a geometric distribution with parameter p. Find a sufficient estimator for p.

The pdf of geometric distribution is: $f(x) = (1-p)^{x-1}p$. We write it in exponential form as,

$$f(x) = \exp\left[x\ln(1-p) + \ln\left(\frac{p}{1-p}\right)\right]$$

We don't have a function for x yet, but that's easy to deal with. Use the fact that $\ln(x^0) = 0$, we can construct a pdf just as the exponential theorem defines.

$$f(x) = \exp\left[x\ln(1-p) + \ln\left(x^0\right) + \ln\left(\frac{p}{1-p}\right)\right]$$

4 Two or More Parameters

The Fisher-Neyman theorem can easily be extended to accommodate two or more parameters. We give the equation below. The proof is pretty much the same as for single parameter, so is omitted.

Fisher-Neyman for Two Parameters Let $X_1, X_2, \ldots, X_n$ be a random sample from distribution with a pdf which depends on two parameters θ_1, θ_2. Then $\hat{\theta}_1 = h_1(x_1, x_2, \ldots, x_n)$ and $\hat{\theta}_2 = h_2(x_1, x_2, \ldots, x_n)$ are sufficient statistics if and only if:

$$L(\theta_1, \theta_2) = \phi[h_1(x_1, x_2, \ldots, x_n), h_2(x_1, x_2, \ldots, x_n); \theta_1, \theta_2] \cdot b(x_1, x_2, \ldots, x_n)$$

Exponential Criterion Let $X_1, X_2, \ldots, X_n$ be a random sample from distribution with a pdf of the exponential form:

$$f(x; \theta_1, \theta_2) = \exp[K_1(x)p_1(\theta_1, \theta_2) + K_2(x)p_2(\theta_1, \theta_2) + S(x) + q(\theta_1, \theta_2)]$$

with a variable x does not depend on parameters. Then, the statistics $\hat{\theta}_1 = \sum_{i=1}^{n} K_1(X_i)$ and $\hat{\theta}_2 = \sum_{i=1}^{n} K_2(X_i)$ are jointly sufficient for true θ_1, θ_2.